DOCUMENT RESUME

ED 428 931 RC 021 96:

AUTHOR Miyashita, Mizuki; Moll, Laura A.
TITLE Enhancing Language Material Availability Using Computers.
PUB DATE 1999-00-00
NOTE 5p.; Chapter 9 in: "Revitalizing Indigenous Languages. Papers presented at the Annual Stabilizing Indigenous Languages Symposium (5th, Louisville, KY, May 15-16, 1998)"; see RC 021 955.
AVAILABLE FROM Web site: http://jan.ucc.nau.edu/~jar/RIL_Contents.html
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Oriented Programs; *Dictionaries; *Information Processing; *Language Maintenance; *Material Development; *Papago; *Second Language Instruction; Tohono O Odham People
IDENTIFIERS *Web Sites

ABSTRACT

This paper describes the use of computer technology to produce an updated online Tohono O'odham dictionary. Spoken in southern Arizona and northern Mexico, Tohono O'odham (formerly Papago) and its close relative Akimel O'odham (Pima) had a total of about 25,000 speakers in 1988. Although the language is taught to school children through community and formal education, these language stabilization efforts have been ineffective due to limited availability of materials and qualified teachers and the fact that Tohono O'odham is not spoken in most homes. Two Tohono O'odham dictionaries are currently used by language learners and scholars--the Mathiot dictionary and the Saxton, Saxton, and Enos dictionary. Each has weaknesses and strengths; neither is written in the Alvarez-Hale writing system, the official orthography of the Tohono O'odham Nation. The Tohono O'odham Dictionary Working Group aims to combine these dictionaries, using the Alvarez-Hale orthography. In a preliminary project, the more comprehensive Mathiot dictionary, which is out of print, is being made accessible online. Project steps include gaining permission of the copyright holder; scanning the text and making formatting changes to regularize the text; proofreading and correcting main entries; generating a Tohono O'odham spell-checking program to correct the rest of the text; converting to the Alvarez-Hale orthography; and creating a searchable Web page. Project benefits and future related projects are discussed. (SV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 428 931

Enhancing Language Material Availability Using Computers

Mizuki Miyashita and Laura A. Moll

This paper describes the authors' use of computer technology to produce an updated online Tohono O'odham dictionary. Access to endangered language materials can be an important factor in the revitalization of languages. The authors describe issues encountered in converting an out-of-print dictionary into a widely available, computer-readable resource, detail solutions that have been developed, and suggest that this process is transferable to materials in other languages.

Jon Allan
Reyhner

As computer technology develops and becomes more popular, it is being introduced in many Native American communities, primarily through schools. Furthermore, computer literacy skills are becoming necessary for survival in the modern workplace. At the same time that use of this new technology is becoming widespread, indigenous languages are being spoken less and less. Language revitalization efforts can benefit from more active use of computer resources.

The project described in this paper directly uses computer technology to make native language material available more widely in order to allow its use for language learning and research. This project provides the Mathiot dictionary, an out-of-print Tohono O'odham to English dictionary, in an online format. It also converts the dictionary to a Tohono O'odham to English and English to Tohono O'odham dictionary in the process. We are putting this dictionary online because it is currently unavailable to most people since it is out-of-print, and the availability of materials for language learning and literacy is very important, especially for an endangered language. Additionally, we change the orthography used in the dictionary to the Alvarez-Hale writing system, which is the official orthography of the Tohono O'odham Nation (Zepeda, 1983). In this way, we encourage more consistent use of the official orthography.

We begin by describing the Tohono O'odham language community. This is followed by a discussion of the existing Tohono O'odham dictionaries. We suggest that use of computer technology is advantageous for language stabilization and describe the process of converting a print dictionary into a searchable online dictionary. We finish with a summary of this project and future related projects.

Language community background

Tohono O'odham (formerly Papago) is spoken in Southern Arizona and Northern Mexico. Tohono O'odham and its very close relative, Akimel O'odham (Pima), had a combined total of approximately 25,000 speakers in 1988 (Fitzgerald, 1997). While many adults speak the language, few children are learning it as their first language.

Through the tribal community and formal education, the language is taught to school children. However, this education does not have a great impact on

RC 021964

BEST COPY AVAILABLE

language stabilization and revitalization owing both to the limited availability of materials and qualified teachers and to the fact that Tohono O'odham is not being spoken in most homes. The Tohono O'odham Tribal Policy encourages the use of the language within the community (Zepeda, 1990). However, the tribe cannot enforce language use among tribal members, and English is commonly used by Tohono O'odham people.

Dictionary resources for Tohono O'odham

There are two dictionaries of Tohono O'odham currently used by language learners and scholars. Both are useful in different ways, but neither is written in the Alvarez-Hale system. The Saxton, Saxton, and Enos dictionary (1983) is most commonly used in Tohono O'odham language courses. It is useful in that it has both Tohono O'odham to English and English to Tohono O'odham entries. However, it contains a limited number of entries. Additionally, the entries do not include much grammatical information or any example sentences.

The Mathiot dictionary (1973) is much more comprehensive than the Saxton/Saxton/Enos dictionary. This dictionary gives more than 11,000 entries, which include detailed grammatical information and example sentences. However, it gives entries only from Tohono O'odham to English and is out-of-print.

Both of these dictionaries are good resources for the Tohono O'odham language community. Each of them has weaknesses that are complemented by the strengths of the other. A combination of these dictionaries, in the Alvarez-Hale writing system, would be ideal.

The Tohono O'odham Dictionary Working Group is working to create just such a dictionary. This is a tribal group that is concerned with stabilizing the language. The group envisions the dictionary as a five-year project and is solidifying a plan for its creation at this time. This group will use entries from the Saxton/Saxton/Enos and Mathiot dictionaries as a foundation, but they intend to type this information by hand. Our project will allow us to provide the dictionary working group with computer disks containing the Mathiot dictionary information. This will save the group much time and effort.

Advantages of using computer technology

There are several ways that the out-of-print Mathiot dictionary could be made available, and there are many advantages to making the dictionary accessible online. In that format, it has the widest potential availability because people can use it without having to buy it. In addition, an online dictionary allows richer searches than a printed dictionary, which is useful for language learners and language researchers. Computerization of the information in the dictionary also allows for easy conversion from Tohono O'odham to English entries to English to Tohono O'odham entries. In addition, an online dictionary of the Tohono O'odham language provides a higher profile for the Tohono O'odham Nation.

Process

The main parts of the process of putting an out-of-print dictionary online are gaining permission of copyright holder, scanning the text, editing the text,

and creating the online dictionary. The first issue to consider in using previously printed material is the copyright. Copyrights on dictionaries are unusual because the entries in the dictionary are not copyrightable as the words themselves are facts, and facts can not be copyrighted. However, the formatting, example sentences, and instructions for dictionary use are created by the author, so they are copyrightable. Since we use the example sentences and grammatical information included in the Mathiot dictionary, we must obtain permission from the copyright holder in order to make this information publicly available.

We are scanning the 864 pages of the dictionary because we estimated that the character recognition of scanning reaches an accuracy rate of 75%. Thus, scanning the dictionary is much faster than typing the entire text. There are several steps involved in the process of scanning. The first is to scan each page, which is like taking a picture of the page and storing it in a computer. Next, we use an Optical Character Recognition (OCR) program to change the picture to characters that can be worked with in a word processing program. The third step is to paste the scanned data into a word processing document. The scanned characters may appear in several different formats, which may also differ from the original text. Therefore, the final step in the scanning process is to make formatting changes in order to regularize the font size and to remove text that is in boldface, italics, and so forth. This entire procedure takes approximately three minutes per page, from book to word processing file.

After completing the scanning process, the dictionary entries are proofread because, as mentioned earlier, the scanning accuracy is about 75%, meaning that 25% of the scanned text is incorrect. In order to obtain a faithful copy of the dictionary, we begin by correcting only the main entries. This is because each O'odham word in the dictionary text needs to be represented by a main entry. First we make global corrections to the main entries using a Perl computer program, and then we manually check the entries in the word processing document because some incorrectly scanned characters involve one-to-many correspondences and others involve special characters, neither of which can be globally corrected using Perl.

Following the correction of the main entries, we generate a Tohono O'odham spell-checking program from these entries and use that program to correct the spelling in the rest of the O'odham text. At this point, we have all the Mathiot dictionary entries in a word processing document. We convert the text to the Alvarez-Hale orthography, and then we are ready to create a web page containing the Mathiot dictionary in a computer-searchable form. Eventually our temporary web page (currently at http://w3.arizona.edu/~ling/mh/lmmm/to.html) will provide all the following features:

1. A space for the user to enter a word in Tohono O'odham or English;
2. A Perl program that returns the meaning(s) of the entered word in the other language;
3. Grammatical information for the Tohono O'odham entries;
4. Example sentences in both languages;

5. Searches by first part of word, last part of word, whole word, or part of word;
6. Suggestion of closely spelled entries if the searched-for entry is not in the dictionary;
7. Links to other O'odham pages (language, culture, etc.); and
8. A description of the steps used to create this online dictionary.

Conclusion

In this paper we discuss how to make out-of-print materials available using computer technology and the resulting beneficial results. One specific result of this project is that it makes this language learning and research information widely available. Also, we are able to provide the Mathiot dictionary information to the community Tohono O'odham Dictionary Working Group in a computerized format. The project makes the dictionary information available in several formats on disks for future purposes. Additionally, there is a comprehensive dictionary available in the Alvarez-Hale writing system, which helps literacy development and encourages consistency in orthography. Finally, the process itself is transferable to dictionaries and other texts of various languages.

There are several related projects that we plan for the future. Once the dictionary is completed, we plan to offer tutorials on its use for students, teachers, and other members of the Tohono O'odham community. The tutorials will include basic computer skills, such as how to use a mouse or how to get online, if needed. We will also request feedback on its ease of use and utility. Finally, we plan to support other language groups with similar projects through a description of the process (on a web page) and direct help.

Note: The authors wish to thank Michael Hammond, Terry Langendoen, Madeleine Mathiot, Delbert Ortiz, Carrie Russell, and Ofelia Zepeda for their support of this project. The authors can be reached at mizuki@u.arizona.edu or mollmoll@u.arizona.edu

References

Fitzgerald, Colleen M. (1997). *O'odham rhythms*. Unpublished doctoral dissertation, University of Arizona.

Mathiot, Madeleine. (1973). *A dictionary of Papago usage*. Bloomington: Indiana University.

Saxton, Dean; Saxton, Lucille; & Enos, Susie. (1983). *Dictionary: Papago/Pima—English, O'othham—Mil-gahn; English—Papago/Pima, Mil-gahn—O'othham* (2nd Ed.). Tucson: University of Arizona.

Zepeda, Ofelia. (1983). *A Papago grammar*. Tucson: University of Arizona.

Zepeda, Ofelia. (1990). American Indian language policy. In K.L. Adams & D.T. Brink (Eds.), *Perspectives on Official English: The campaign for English as the official language of the USA* (pp. 247-256). Berlin: Mouton de Gruyter.

BEST COPY AVAILABLE